# Making music with images: interactive audiovisual performance systems for the deaf

M Grierson

Department of Music, Goldsmiths College,
Lewisham Way, New Cross, London, UK

*m.grierson@gold.ac.uk*

*www.goldsmiths.ac.uk/ems*

## ABSTRACT

This paper describes the technical and aesthetic approach utilised for the development of an interactive audiovisual performance system designed specifically for use by children with multiple learning difficulties, including deafness and autism. Sound is transformed in real-time through the implementation of a Fast Fourier Transform (FFT) and translated into a moving image. This image is adapted so that relevant information can be understood and manipulated visually in real-time. Finally, the image is turned back into sound with only minimal delay. The translation process is based on research in computer music, neuroscience, perception and abstract film studies, supported by the Arts and Humanities Research Council. The system has been developed through collaboration with the Sonic Arts Network, Whitefields Special Needs School, and the South Bank Centre, specifically for a project led by Duncan Chapman with the London Philharmonic Orchestra. The system has now been made available for free by the Sonic Arts Network.

## 1. INTRODUCTION

Through collaboration with the Sonic Arts Network, Whitefields Special Needs School, the South Bank Centre and the London Philharmonic Orchestra, a software system has been developed that visualises sound in real-time in a way that allows hearing-impaired individuals to interact with a specifically designed, experiential representation of sound.

Deaf children with learning disabilities can interact with the system through the use of a Nintendo Wiiremote, which provides additional vibration-based feedback when the audio signal peaks over a set amplitude. Users can load and play back sounds from the remote, whilst using the accelerometer to make real-time adjustments to the pitch, pan position and volume. In addition, the system accepts live input, giving users the ability to interact with musicians in a meaningful way, either by changing the nature of the sound in real-time, or by recording and playing back segments of live audio. Beyond this, the musical output of the software can be set to a number of transposable scale systems. This allows for a wide range of musical interaction.

Musical interaction is conducted through the alteration of the visualisation itself. Through learning to control a number of specific gestures, the users can alter the appearance of the visualisation instantaneously. The visualisation is then converted back into audible sound, perceptually occurring at the same time as the original sonic material. The sound and music is effectively 'translated' into useful visual information that can be sensibly interpreted and changed in real-time by users who are unable to easily hear and react to the sound itself. Most significantly, the system allows for users to become more aware of their own sound making in an informative, detailed, and immediate way, providing useful information that could aid in the development of a user's relationship with sounds. It is hoped that this system will have genuine therapeutic benefits for those without normal hearing, particularly in relation to the development of voice production.


©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

## 2. THE VISUAL REPRESENTATION OF SOUND

### *2.1 Existing Standards for Sound Visualisation*

Several standardised methods for visual representation of sound are used everyday by musicians, engineers, and others that need them. These methods can be loosely divided into three types: Symbolic representation, average level indication, and spectral analysis.

Symbolic representations, such as musical scores, graphic notations and markup languages display information using an agreed set of symbols. These forms of representation only convey a certain amount of information, given a specific set of limitations. Most importantly, their relationship to sound is not indexical (in the sense of Peirce (in Hartshorne et al, 1931-58)) - i.e. the sounds and images are not in 1:1 relationships.

Level indicators such as PPM (Peak Program Meters), VU (Volume Unit) and dB (decibel) meters display changes in the average amplitude of sound signals. These are less like symbolic representations in that their relationship to sound is indexical. However, the amount of sonic information contained in these displays is limited. Information is only apparent in one dimension – amplitude. For example, it can be considered impossible to discern pitch relationships by viewing a level meter.

Spectral analysis is a robust method for sound visualisation. Spectrographs & sonograms function by decomposing sounds into a number of component sinusoids, and plotting their amplitudes on a graph. This is usually performed using a Fourier Transform, although other (related) methods exist.

In the case of providing visualisations for those with disabilities for the purpose of interaction, all of these methods could be improved, either lacking in sufficient information, or being difficult to understand. Symbolic representations often contain little textural or experiential information – that is to say, symbolic methods can be used to describe how a sound changes over time, and what pitches it might contain, but not what it is like to hear the sound. Level meters provide some experiential information, but this is heavily limited. Spectral methods provide a great deal of information with respect to the power spectrum of sounds – but this information is not presented in a way which makes it easily understandable, nor experientially relevant.

### *2.2 Audio Visualisation in Experimental Film*

Within the discipline of experimental film there have been many highly regarded attempts to visualise sound and music in ways that are experientially relevant. Key examples of this practice include *Rhythmus 21* (Richter, 1921), *Diagonal Symphony* (Eggeling, 1922-24), *Studie No. 6,* (Fischinger, 1930), *Allegro* (McLaren, 1939) and *A Colour Box* (Lye, 1935). To some extent, information about sound is translated into moment-to-moment changes in shape and texture in an attempt to echo the experience of sound. An excellent example of this practice can be found in the work of John and James Whitney, notably *5 Abstract Film Exercises,* (Whitney, 1944-45). These practices have been well documented by scholars such as Al Rees (1999), P. Adams Sitney (1974), and Malcolm Le Grice (1977), among many others.

Importantly, these relationships can easily be shown to be arbitrary, containing little or no musical or detailed sonic information (similar to symbolic and level indication methods, whilst sometimes containing textural information). Despite this, what is key about this practice is that it emerges from the idea that elements of our sonic experience – specifically in terms of motion - can be experienced visually. As Sitney mentioned on the occasion of film-maker Stan Brakhage's death (Sitney, 2003), Brakhage was committed to the concept of 'moving visual thinking', the idea that internal cognitive visual processes could be made visible. Even though Brakhage had often discussed his practice as being related to music, there is a tension here between the desire to explore purely visual processes, and the concept that vision can be similar to musical experience. However, through this tension rises potential new solutions to the problem of how to effectively represent sound for the purposes of audiovisual interaction, particularly in situations where participants have a very different experience of sound.


©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

## 3. NEUROSCIENTIFIC EVIDENCE FOR THE EFFECTIVENESS OF AUDIOVISUAL CONGRUENCE

Evidence supporting the view of experimental film-makers that the experience of sounds and music can be fundamentally linked with the experience of moving images continues to grow. Key examples demonstrating the stimulation of multisensory cells in the Superior Colliculus by audiovisual interaction include the McGurk effect (McGurk et al, 1976), and the double-flash illusion (Shams et al, 2000). In addition, evidence that seemingly arbitrary links between sound and image can improve event perception is emerging (Effenberg, 2001). The idea that a visualisation based on perceptual congruence may aid in translating the experience of sound for non-hearing populations has not been effectively tested, and there may be several problems with this view – especially considering that the precise nature of audiovisual relationships often appears arbitrary and is not always well understood.

Our research does not necessarily claim to have scientifically demonstrated a definitive link between a particular type of visualisation and a particular type of sound. Instead, we offer a technically useful and experientially relevant method for translating sound to an image and back again that may prove to be effective for enhancing musical and sonic interaction, particularly (but not exclusively) for the deaf and hard of hearing. Most importantly, we have chosen a visualisation technique directly inspired by experimental film practice, and neuroscientific enquiry that directly relates to the structure of the visual cortex (Bressloff et al, 2001).

## 4. METHOD

This project combines experiential approaches that share similarities with symbolic and indexical forms of audio visualisation (such as those found in experimental film making practice), with filtered spectral analysis and signal processing methods (FFT and related processes). This introduces a level of information not commonly found in similar approaches. Importantly, not all the available spectral information is used in the visualisation. Attempts have been made to reduce the amount of information so that only the most important elements are retained. Once the sound has been translated into image, it can be manipulated in real-time by a user. Importantly, the manipulated image can then be turned back into sound. This process occurs quickly enough that the users remain unaware of any significant time delay. Crucially, it is the real-time editing of this image which results in the perceptually relevant changes in the sound.

### *4.1 System Design and Aesthetics*

The project was realised using Cycling 74's Max/MSP/Jitter, based on Miller S Puckette's Patching environment (Puckette, 1988). The sound is converted into a real-time spectrogram (see Figure 1). This spectrogram is loaded into a one-dimensional 32 bit floating point array, 2048 bins in size (see Figure 2.). It is then modified to appear as a set of two-dimensional concentric rings. In order to preserve processor load, the calculation from one to two dimensions is done using the computer's video hardware. This is achieved using a texture mapping process, with the one dimensional array being stretched across an OpenGL Surface. The OpenGL surface can be manipulated using a third dimension, to alter the appearance of the concentric ring formation with very limited CPU overhead.

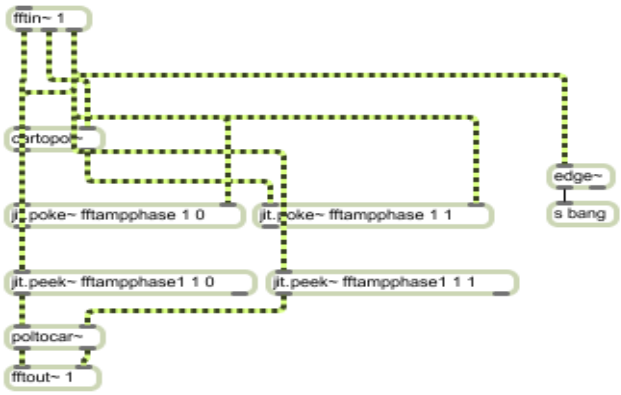


**Figure 1.** *Max/MSP Patch showing an FFT process being sent to a video buffer.*


©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

This results in an additional speed gain, and is useful for filtering the visual information without altering the material in the video buffer itself. This means that the image can be made more perceptually relevant without losing valuable data that is needed in order to re-synthesise the signal. Most importantly, all data manipulation continues to be performed on the one dimensional array, whilst being displayed in three dimensions.


**Figure 2.** *The FFT data expressed as one dimensional video information*

The concentric rings are spaced non-linearly so as to more accurately mirror human perception of audible frequencies. The exact spacing of the rings is achieved through the application of a logarithmic function on the visualisation of the 2048 point FFT spectrum. This function is applied by mapping the information onto a 3D OpenGL sphere (with non-uniform dimensions). This results in some spectral information loss with respect to the display. This is acceptable for two reasons. First, the spectral information is still maintained in the one dimensional array. The shifting in pitch appears relative, whereas if the display were linear, the shifting would appear to spread out as the fundamental frequency rose. Secondly, it is not required that the display be scientifically accurate, just that it remains perceptually relevant, whilst still resulting in an accurate re-translation back into sound.

### *4.2 The Concentric Ring Formation as a Form Constant*

There is a precedent for the representation of sound and music in a concentric ring formation that is clearly demonstrated in the field of early 20th century film making and animation, such as the visual music tradition. Oskar Fischinger's silent film *Spiralen* (Fischinger, 1924) uses a number of concentric and spiral formations in what can be described as an early attempt at producing a visual metaphor for the experience of sound. In addition, this type of representation is echoed in the work of computer graphics and visual music pioneer, John Whitney. Although neither of these artists exclusively used the concentric formation, the iterative tunnel effect which it produces fits with the forms that continually appear in this canon.

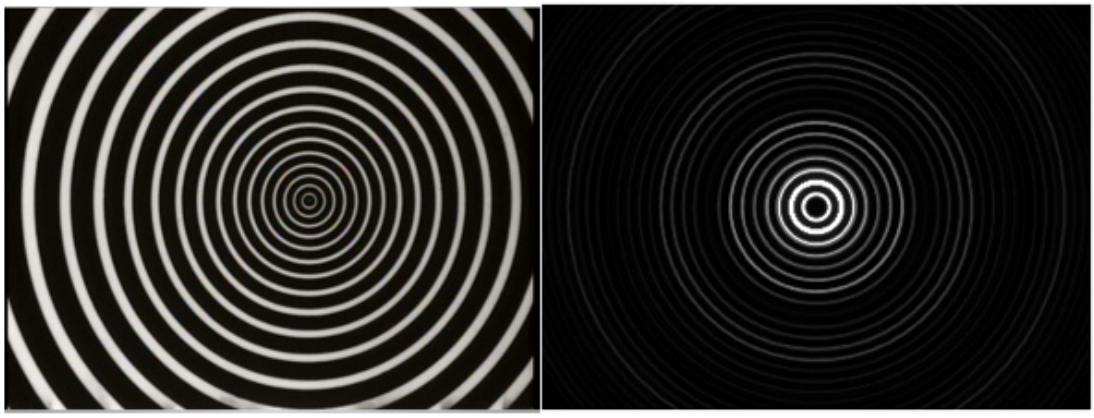

**Figure 3.** *Still image from Spiralen (1924) by Oskar Fischinger (left), and a screengrab from the audiovisualisation software (right).*


©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

Further reasoning behind this choice of visualisation comes from studies in neuroscience and perception. In Bresslof, P., et al (2001) a number of visual patterns are identified as persisting in various forms of visual art and experience generally as a result of their direct mathematical relationship with the structure of the visual cortex. These visual configurations are roughly divided into four categories - tunnels and funnels, spirals, lattices (which include honeycombs and triangles), and cobwebs. The formation described here fits more than one of these categories at various times depending on the nature of the sound being analysed / re-synthesised. However, due to the perspective illusion that it evokes, it most closely fits within Bresslof et al's 'Tunnels and Funnels' category.

These precedents provide theoretical weight to the idea that a very specific type of concentric visualisation may be effective with respect to the goals of the project. The Concentric ring display format was chosen in an attempt to encourage the users, most of whom were autistic in this case, to focus on the image as a focal point of their experience, and to become sensitive to the complex and potentially delicate variations in the sound that result from the user altering the image through their own interaction. It is also being employed to hold their attention, and to give them an indication of the morphology of the sound and music that other, able bodied participants are experiencing.

Sonic transformations are performed by moving the FFT bins in both directions. These bin shifts are logarithmic – frequency relationships remain coherent throughout musical octaves. In addition, phase information is rotated, enabling the system to operate as a phase vocoder. As such it provides smooth real-time pitch correction, and can be used as a high quality real-time time-stretching device. In addition, the system allows for spectral freezing and Fourier synthesis through manipulation of frozen spectral fragments.

## 5. RESULTS

Working alongside the Sonic Arts Network, Whitefields school, and in collaboration with experienced workshop leader, Duncan Chapman, various versions of the software were tested with a number of users. Feedback was very positive with respect to the effectiveness of the visualisation. Immediately it became apparent that the software gave children with multiple learning disabilities including deafness the opportunity to appreciate the way their own sound making was experienced by others. Through comparing the experience of 'video-listening' to pre-recorded and live sound material with the experience of their own interaction and sound making activity, users were able to more effectively appreciate the effects of sound in the world around them. As such, the tool was considered immediately useful in (re)habilitation, giving non-hearing users a simple way of having meaningful relationships with sound and music.

Two significant problems were identified and overcome throughout the testing period. First, the system allowed for non-hearing individuals to synthesise sounds through a visualised form of Fourier synthesis. It was found that although this could be entertaining, it often produced a wide variety of unpleasant sounds. These functions were disabled, although they may be added again later as the control functions become increasingly refined. In addition, it was noted that the graphical interface could be developed to enable teachers and students to use the system unaided by the workshop leader. As a result of this process, a GUI was designed through collaboration with an information designer with knowledge of teaching environments (Figure 4).

A later session, organised and filmed by BBC online technology at the Frank Barnes School for the Deaf, showed that the system could be successfully used by students and staff with little difficulty. In addition, this session demonstrated the system's capacity for use in voice production training. Through focussing on the image, students were able to respond to very small changes in the sounds that they were producing. As students began producing sounds, they appeared immediately aware of the way changes in a sound's tone and quality were being represented. This seemed successful in that students were able to easily begin to produce similar sounds by watching the variations on the display, and attempting to repeat sounds reliably. Staff at the Frank Barnes school were confident that the tool could be of great benefit to many with respect to voice training. As a result, plans are being made to deploy the software throughout a network of deaf schools in the UK.

Finally, Duncan Chapman has developed a set of workshop exercises that can be used as a starting point for those wishing to use the system. These cover three main areas: voice production, recording and playback, and musical interaction with other performers. The workshop exercises are distributed as part of the download package, which can be found at www.sonicartsnetwork.org.


©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

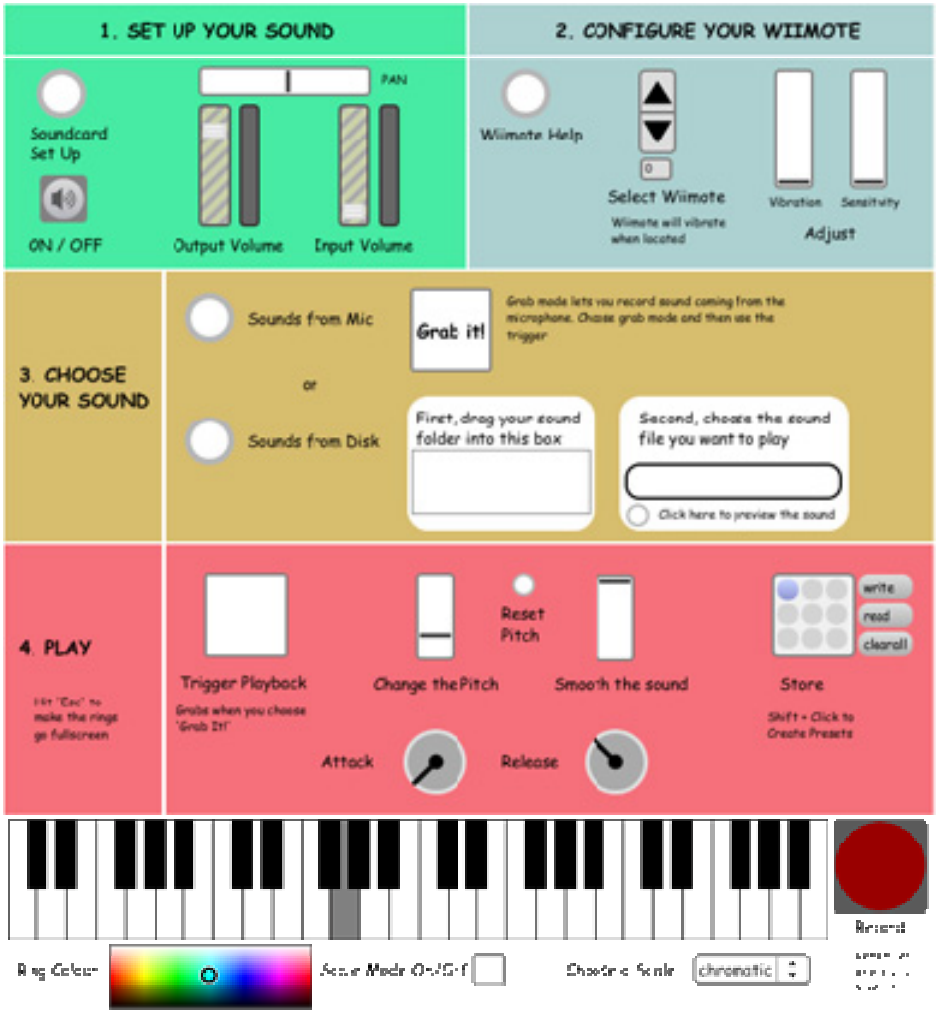


**Figure 4.** *The Graphical User Interface, designed in consultation with information designer Myah Chun.*

## 6. CONCLUSION

This system has been used in number of collaborative workshops and performances at the South Bank Centre, London, and at the Frank Barnes School for the Deaf, Camden, London. Some of these sessions featured members of the London Philharmonic Orchestra working with disabled students from Whitefields school. These events were documented by the Sonic Arts Network, and the BBC. In addition, several compositions were directed by Duncan Chapman. The system has been improved to allow ease of use for both hearing and non-hearing players of varying ability so that they may collaborate in performance in ways that are immediate and inherently musical. Beyond this, the system has shown genuine potential for use as a therapeutic tool.

The software was made publicly available free of charge in 2008 by the Sonic Arts Network. In addition, some source code will be released as part of the AHRC funded project, Cognitive and Structural Approaches to Contemporary Computer Aided Audiovisual Composition. This project was demonstrated as part of a BBC technology article, due for publication in Summer 2008.

**Acknowledgements:** The staff and students of Whitefields School for the disabled, London.


©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6

## 7. REFERENCES


P Adams Sitney (1974), *Visionary Film: The American Avant-Garde,* 1943-78 New York: Oxford University Press.

P Adams Sitney (2003), Jusqu'a son dernier souffle, *Cahiers du cinema* 578, pp. 50-1, Editions de l'Etoile.

P C Bressloff, J D Cowan, M Golubitsky, P J Thomas and M C Weiner (2001), What Geometric Visual Hallucinations Tell Us about the Visual Cortex, *Philosophical Transactions of the Royal Society* B, 356, 2001.

A O Effenberg (2001), Multimodal Convergent Information Enhances Perception Accuracy of Human Movement Patterns, *Proc. 6th Ann. Congress of the European College of Sports Science (ECSS)*, Sport und Buch Strauss, p.122.

V Eggeling (1922-24), *Diagonal Symphony.*

O Fischinger (1924), *Spiralen*.

O Fischinger (1930), *Studie No. 6.*

M Le Grice, (1977) *Abstract Film and Beyond*, MIT

L Lye (1935), *A Colour Box*.

H McGurk and J MacDonald (1976) Hearing lips and seeing voices, *Nature,* Vol 264(5588), pp. 746–748.

N McLaren (1939), *Allegro*.

C S Peirce (1931-58), *Peirce, Charles Sanders : Collected Writings,* (C Hartshorne, P Weiss, & A W Burks, Eds), Cambridge, MA: Harvard University Press.

M Puckette (1988), The Patcher, *Proceedings, ICMC, International Computer Music Association*, San Francisco, pp. 420-429, 1988.

A L Rees (1999), *A History of Experimental Film and Video*, BFI.

H Richter (1921), *Rhythmus* 21.

L Shams, Y Kamitani and Y Shimojo (2000), What you see is what you hear, *Nature*, Vol 408, pp. 788.

J Whitney and J Whitney (1944-45), *Five Abstract Film Exercises.*




©2008 ICDVRAT/University of Reading, UK; ISBN 07 049 15 00 6